

What if there were an African alliance for meaningful access to Intermediaries' data holdings?

Draft document for consultation at the Africa IGF,
Abuja, 18 September 2023

Background

Research ICT Africa (RIA) is a think-tank with a long track record in generating knowledge on digital matters. Amongst its current projects is one in co-operation with the Action Coalition for Meaningful Transparency (ACT), of which RIA is a member. This investigates platform transparency as pertains to African stakeholders. Similar research initiatives are taking place in Asia and Latin America. Comments are invited on this draft and may be sent to: ATD@researchictafrica.net

Introduction:

Access to data, and the use of such data, is a significant variable for progress in Africa towards the sustainable development goals and Agenda 2063. This is highlighted most recently by the rise of generative AI, trained on large data sets in which African languages for instance have minimal presence. Among the biggest holders of African-generated data at scale and collected over substantial time periods are electronic communications Intermediaries – i.e. internet and telecoms companies. Many of these entities have made general commitments to transparency and have taken some steps towards openness in some regions. However, the current African reality is one of limited access and uptake of Intermediaries' data holdings. But the potential exists to change this.

The research underpinning this document¹ has revealed a general lack of awareness of the significance of Intermediaries' data holdings, but also some appetite for increasing access and use. This means there is interest in the “demand” side that includes - but also goes beyond - what is currently availed in transparency reports or the [existing access possibilities](#) as currently afforded by the companies. This attention is evident amongst academics (particularly those advocating Open Science), as well as data journalists, NGOs with interests in various subjects (eg. monitoring disinformation, gender equality, climate-related online content, fact checkers), regulatory instances (eg. Electoral management bodies) and data-based start-ups. Currently, substantive access to Intermediaries' data holdings requires going through costly brokers (which are too expensive for most African stakeholders) and otherwise by informal scraping of public-facing data on the platforms. A few commercial African entities have partnerships with Intermediaries which entail data-co-operation between the two sides.

Global momentum for access to data:

In a context of increasing transparency by Intermediary companies, it is evident that numerous calls have mounted for this to extend into meaningful access to their data holdings.

“In order to study any of the pathologies attributed to the transformation of the digital media ecosystem, social scientists need access to platform-controlled data as to “who” saw or engaged with “what” “when.” That is, scientists need to understand how and when certain people, and the population at large in different countries, interact with new media and what the consequences of those interactions are. Even when the platforms have promised to make available data for independent academic research, those promises have often gone unfulfilled.”

¹ Key Informant Interviews with 20 experts and data-related practitioners in diverse sectors from around sub-Saharan Africa.

“Public authorities must compel major Internet platforms to provide independent parties with meaningful data about the impact social media has on democracy.

Protecting Electoral Integrity in the Digital Age. The Report of the Kofi Annan Commission on Elections and Democracy in the Digital Age. January 2020

Many calls have been framed in terms of understanding and addressing problems attributed to Intermediaries, but there is also the potential for public-interest data access to contribute to developmental objectives such as health interventions or public planning. This interest has led to practical opening up of data under certain conditions. Both voluntary and compulsory arrangements around access are emerging which seek to prevent entirely self-serving corporate controls on the one hand, and avoid Cambridge-Analytica-style privacy abuses on the other.

For example, on the data “supply” side, the large corporate members of the [Global Network Initiative](#) (including major international platforms and telcos) have some voluntary momentum in assessing and addressing obstacles to data access through the [Action Coalition for Meaningful Transparency](#). Internationally, there are increasing cases of key Intermediaries voluntarily opening up in selected areas. Examples of this include [Meta’s provision of access for 17 US academics](#) and two research surveys through [a research interface](#) during the 2020 elections in the US.² This raises the question of: if there, why not also in African countries?

Legal developments requiring data access include draft legislation in the US ([The Platform Accountability and Transparency Act](#)), covering both university-affiliated researchers as well as nonprofit organizations’ researchers. Data access proposals would be vetted by the US National Science Foundation with possible involvement of the Federal Trade Commission.

Many companies are also preparing to deal with compulsory access directives under [European Union regulations](#), which will be mediated by the national Digital Services Coordinator (DSC) of Establishment to be established as EU member states’ regulatory bodies. The major platforms have committed to research access via an advisory “independent, third-party intermediary body” under the EU’s revised [Code of Practice on Disinformation](#) in June 2022. This Code includes provisions that companies will not prohibit, discourage or take “adversarial action” against public-interest good faith research into disinformation on their platforms.³ It also commits signatory platforms to provide fact-checkers with information and appropriate interfaces to help them quantify the impact of fact-checked content, such as actions taken based on that content, impressions, clicks, or interactions.

Big tech companies have further agreed to a [draft code of conduct](#) on researcher access to data, and the need for an independent mediatory body to adjudicate data access, as authorised by the EU’s General Data Protection Regulation (GDPR). This draft code, adopted after extensive negotiations with the [European Digital Media Observatory](#), is accompanied by a model data-sharing agreement that would ensure privacy compliance.

² Four studies published in Open Access in Nature and Science journals in mid-2023 covered: How do social media feed algorithms affect attitudes and behavior in an election campaign?; Asymmetric Ideological Segregation in Exposure to Political News on Facebook; Reshares on social media amplify political news but do not detectably affect beliefs or opinions; Like-minded sources on Facebook are prevalent but not polarizing.

³ See footnote 9 below.

Generalised access via regulatory arrangements is not likely to happen in individual African countries in the short term. However, the developments outside the continent could have a positive spill-over impact in Africa and/or open up the possibility of partnerships between Africans and those abroad (including Africans in the diaspora) who are eligible to benefit from offshore developments.

The interviews for this research unearthed that there are potential prospects to build a virtuous data chain in Africa, with the possibility of moving toward modular experiments around specific use cases and partnerships in public interest. While these prospects would need to deal with important issues such as data security, quality, retention, ownership and standards, this brief focuses only on access.

The value of access to data in Africa

The African Union in 2022 agreed a [Data Policy Framework](#), with an implementation plan being elaborated during 2023. This initiative highlights the value of data in general to researchers, policy makers, NGOs, tech community, branches of government, IT start-ups, NGOs, etc. Data are further key if there is to be evidence-based development policy in Africa, as well as for the monitoring of policy and practice in general, including policy on digital services.

Of relevance to the issue of accessing Intermediary holdings, African countries have made major strides towards improving broad policy ecosystems about data. By mid-2023, the Malabo Convention on Cybersecurity and Data Protection [had entered into force](#) with the required minimum of 15 states ratifying the agreement. [Researchers say](#) that of 55 African countries, 35 have comprehensive data protection laws with 3 having draft laws. A total of [27 states have access to information laws, with 17 draft laws in existence](#) - some of which extend to private sector data holdings. The African Union Commission's [Declaration of Principles for Freedom of Expression in Africa](#) (2019) notes that "Every person has the right to access information held by public bodies and relevant private bodies expeditiously and inexpensively", and that "Every person has the right to access information of private bodies that may assist in the exercise or protection of any right". As UN Members, African countries have signed up to the Sustainable Development Goals (SDGs), where indicator 16.10 calls for "public access to information and fundamental freedoms" and reporting thereon. All these developments and related norms support increasing access to Intermediaries' data.

Start-ups, particularly in the AI space, have interests in platform data for training Large Language Models concerning under-served African languages. Moreover, regulators, planners and officials could profitably use telco ISP data records for administration and service delivery. Further interest in access to platform data comes from academics engaged in training programmes for digital research methods.

While there is still a lot of progress to make in elaborating the value of access to private data holdings, prospects exist for this to become an issue on the African agenda. This is notwithstanding generalised obstacles. On the "demand side", these hurdles include very low awareness of the potential and even of existing access possibilities; highly limited skills; and entrenched digital divide issues (along with divides in gender inequality). There is also a lack of understanding of digital research methods, and a shortage of skills in data-science, as well as challenges in access to software and processing power.⁴ The [African Open Science Platform](#) and a number of [smaller initiatives](#) are seeking to mitigate these matters.

⁴ Support can be found in the USA, an amount of \$50 000 AWS credits have been donated to the University of Michigan's [Institute for Social Research](#) to offset computing costs for their Social Media Archive.

On the “supply” side, one obstacle to expanding access is the generally highly unequal treatment of African data interlocutors as compared to the situation in the USA and the EU. There are also companies’ generic concerns about privacy protection and about damage to their public image, especially through timely and critical journalism arising from data access.⁵ These challenges can be mitigated through arrangements for independence in access, and there are also adverse reputational signals at stake for Intermediary companies if they continue to apply different standards to data access in Africa.

Intermediaries may also be concerned with what they see as giving away a monetary asset which could even advantage commercial rivals. However, such competition is far from a reality within the African market given challenges of capital and skills in the face of dominant foreign players. On the contrary, data access – along with data partnerships – could end up being to economic advantage for the corporates, with start-ups adding value to their foundational services and the wider ecosystem.

The benefits on all sides of having systems and mechanisms for openness and co-operation, including making existing access channels meaningful, surpass the disadvantages in the current situation. Without change, there will be a continuation of ignorance of existing access channels, diverse practices of informal and partial scraping, and expensive access via commercial brokers. A new relationship could also help with the quality of research findings about electronic communications in Africa more broadly, in contrast for example to the production of [sweeping claims based on miniscule data-sets](#).

The urgency of the issue:

In recent times, data access for African users has suffered since the open research API at Twitter, now called X, was replaced with an [expensive payment regime](#). However, one entity interviewed for this research said it had paid a relatively low sum for specific access. Nevertheless, Africans have also experienced indirect limits from X’s stance (eg. such as via use of the botnet programme which relied on the previously open API).

Concerning Meta, access to some data was previously available via the Crowdtangle interface and subsequent [Facebook Open Research & Transparency](#) (FORT) for approved academic researchers and “[others considered on a case by case basis](#)”. However, these access arrangements were largely put on hold and followed up in August 2023 by a company [announcement](#) of a “[Meta Content Library and API](#)”. This facility is said by the company to include data fields [in 180+ languages](#) on Facebook and Instagram.⁶

“Global researchers” are eligible to apply if they are affiliated with [an academic institution or a qualified research institute](#). Applications are reviewed by the [Inter-university Consortium for Political and Social Research \(ICPSR\)](#) at the University of Michigan, US. The data sets entailed in the 17 researcher example

⁵ Journalistic [coverage reflecting badly on Meta](#), using the company’s Crowdtangle interface, [is reported as leading Nick Clegg, Meta vice president of global affairs, to complain](#) that “our own tools are helping journos to consolidate the wrong narrative.”

⁶ Data from the Library can be searched, explored, and processed via a [web-based controlled-access interface](#) which does not allow downloading of data (although research outputs can be exported). This limitation on export also appears to be the case with data for deeper analysis that is possible through the company’s Researcher programme, which is done by approved access to the Content Library’s API. The Library offers search across over 100 data fields from Instagram accounts and posts and Facebook Pages, posts, groups and events, with filtering for language, view count, media type, and content producer.

cited above are being archived by the same Consortium, in which 2 Ghanaian and 6 South African entities are among the 750 mainly North American members.⁷

On [advertising data](#), Meta [also says](#) it offers “granular ad-level targeting information for all social issue, electoral and political ads run across Meta’s technologies since August 2020 in 120+ countries”. This dataset includes “ad targeting criteria, such as age, gender, location, Custom Audience and Lookalike Audience inclusion or exclusion, and other detailed targeting selections (demographics, behaviours, or interests).” The company’s [Ad Library API](#) enables a deeper analysis of ads about social issues, elections or politics, as well as all other ads that deliver to the EU. The access to data (outside the EU) [includes](#) targeting data for social, electoral and political ads (such as location, age, gender, detailed targeting, language, custom audiences, and lookalike audiences); impression range, and the reach percentage of people who saw an ad, by age and gender, and information about the locations where the ad reached; as well as spend.

While Meta’s encrypted messaging system WhatsApp is enormously present in Africa, the service appears so far to be reluctant to give access to its massive meta-data holdings (this applies outside Africa as well). Meta uses this meta-data to add to its advert profiling for Facebook and Instagram. This data set could however be important for crisis detection and management. While WhatsApp can be [a useful tool for research](#), and a [means of data collection](#), this is a different matter to accessing metadata about its use and the flow of particular messaging.⁸ While Google trends and some ad-tech data is available, this is without more granular offerings. Google-owned YouTube has [a researcher program](#), which gives the platforms sole power to screen researchers and research projects and is limited to public data. The data sets available include ratings for a given video, sharings, and marking a video as a favorite, among others. To be approved, researchers must be linked to an accredited university, “have a clear research goal, and intend to publish their findings”. Topics have to be [approved by YouTube](#). Academic researchers in [some 60 countries](#), including Ghana, Kenya, Senegal and South Africa, are eligible to apply for access to the API. At the same time, [the YouTube researcher terms](#) also acknowledge eligibility of “any government or other institution required by law or regulation to have access to Program Data”. Specific information about African uptake of this access could not be found for this research.

TikTok use in Africa [faces issues](#). At any rate, African researchers appear to be generally scraping it in a piecemeal fashion. TikTok currently offers some [API access](#) to researchers at US- and European universities, such as numbers of video followers and likes. It approves proposals and then researchers have to abide by restrictive [terms of service](#). It is not evident if/when access will be extended to Africans.

Moreover, access to a level of data holdings (with some Intermediaries) by university entities as well as NGOs is happening via brokers like [Brandwatch](#), [Meltwater](#) and [Netbase](#). However, this exceeds the budgets of many would-be data users. There is evidence of some telecoms in Africa having agreed to commercial deals around their data, but not about access to public-interest researchers.

⁷ The company says the contracts it signed were the same as “other independent researchers who use our data, which is publicly posted on [Social Science One’s web site](#)”.

⁸ The WhatsApp Business Platform API allows paying businesses to “connect thousands of agents and bots to interact with customers programmatically and manually”, which signals the utility of certain metadata for narrow commercial users. Again, the significance can be inferred (and elaborated) from the affordance whereby this business API can enable a business [to retrieve all conversation](#) with the specific phone number/s they are using, by time period and country.

Meanwhile, data protection authorities in Europe are calling on companies [to prevent](#) data scraping. This, along with X's legal action⁹ against an NGO for allegedly contravening anti-scraping terms of service, chills informal access arrangements. Some African actors doing Open Source Intelligence (OSINT), interviewed for this research, feared that they could become caught in this net.

To surmount the challenges for access to Intermediary data holdings in Africa will require building a coalition of interested parties to engage with each other and with the Intermediaries. Considered individually, the diverse actors on the demand side are fragmented and constrained in terms of approaching Intermediaries, not to mention having limited impact on national and continental policy about Intermediary data access. An African alliance for meaningful access to Intermediaries' data holdings would overcome these limits, and could play a key role in stimulating initiatives and capacities. It could also put on the agenda the possibilities of legitimate and efficient clearing houses to mediate supply-and-demand relationships.

As experience elsewhere shows, joint approaches can yield momentum. This is the case with the European Digital Media Observatory (EDMO), which generated a multistakeholder code of conduct for researcher access to data under the EU General Data Protection Regulation. The USA has seen the formation [of an academic coalition](#) to advocate for the issue. In the same country, the Centre for Democratic Technology [brought together a group of US researchers to brainstorm ideal research agendas](#). The [United Nations Inter-Agency Group on Disinformation](#) has developed generic data asks from platform companies. Under the auspices of 45 countries, the [Observatory on Information and Democracy](#) is working to advance a worldwide call for access to Intermediary data.

Questions for possible follow-up

(responses may be sent to ATD@researchICTAfrica.net):

1. *What topic/s or cases could be proposed for a productive pilot access initiative with public value benefit?*
2. *What Intermediary data sets would be needed for such, and be set out in a research proposal?*
3. *How can non-academic partners (eg. journalists, NGOs, civic tech) be brought into this kind of initiative?*
4. *What independence, transparency, quality control systems & mechanisms could support access?*
5. *What capacity would need to be prepared on the "demand side" for a given pilot initiative?*
6. *What support beyond data access could be requested from the Intermediaries?*
7. *What is the role of AU, regional and national stakeholders in advancing this agenda?*

⁹ X, formally known as Twitter, is not a signatory of the EU's Code of Practice on Disinformation(although it was a signatory in 2016 of the EU's [Code of Conduct on Countering Illegal Hate Speech Online](#)). In mid-2023, it [sued the Center for Countering Digital Hate](#) for a fall-off in advertising. It said the Center "[as a registered user of X, scraped data from X's platform in violation of the express terms of its agreement with X Corp](#)". The Center had published research that no action had been taken against most of a sample of Twitter Blue accounts which the Center had reported for "[tweeting hate](#)."